EFFECT OF MATING STRUCTURE ON VARIATION IN LINKAGE DISEQUILIBRIUM[1]

B. S. WEIR AND W. G. HILL[2]

*Department of Statistics, North Carolina State University,
P. O. Box 5457, Raleigh, North Carolina 27650*

ABSTRACT

Measurement of linkage disequilibrium involves two sampling processes. First, there is the sampling of gametes in the population to form successive generations, and this generates disequilibrium dependent on the effective population size (N_e) and the mating structure. Second, there is sampling of a finite number (n) of individuals to estimate the population disequilibrium. ——Two-locus descent measures are used to describe the mating system and are transformed to disequilibrium moments at the final sampling. Approximate eigenvectors for the transition matrix of descent measures are used to obtain formulae for the variance of the observed disequilibria as a function of N_e, mating structure, n, and linkage or recombination parameter.——The variance of disequilibrium is the same for monoecious populations with or without random selfing and for dioecious populations with random pairing for each progeny. With monogamy, the variance is slightly higher, the proportional difference being greater for unlinked loci.

THERE is now an extensive theory on the generation of linkage disequilibrium by drift in finite populations. Several early studies used haploid models on the assumption that these would be adequate for diploid organisms (*e.g.*, HILL and ROBERTSON 1968). Comparisons of these haploid models with those for monoecious populations with random selfing suggest that there is indeed little difference in the results except for loosely linked loci which, although little disequilibrium occurs between them, account for most pairs of loci in most species. The effects of different mating systems on moments of disequilibrium have not been adequately discussed, however, nor has the mating system been related to moments of the within- and between-gametes disequilibria of COCKERHAM and WEIR (1977).

A convenient framework on which to build a theory of linkage disequilibrium is that of two-locus descent measures. For populations initially in linkage equilibrium, only those measures concerned with identity by descent are needed, and a number of transition matrices for these measures in various mating systems have been given by WEIR and COCKERHAM (1969) and WEIR, AVERY and

HILL (1980). Variability in linkage disequilibrium can be obtained from these matrices by suitable transformations.

The observed variation in disequilibrium follows from the sampling processes implied in the measurement of the coefficients. First, there is the sampling of gametes to form successive generations. This between-line variation is controlled by the mating system and is dependent on effective size, N_e. Second, there is the sampling of individuals to be scored, from which estimates are constructed. This is a within-line variation and is dependent on sample size, n.

<div align="center">ANALYSIS</div>

Descent measures: Following WEIR, AVERY and HILL (1980), the 10 two-locus descent measures needed in this analysis are listed in Table 1. They involve genes located in two, three or four gametes and are referred to as digametic, trigametic or quadrigametic recipient measures, respectively. Each measure is a joint probability that genes a and a' at one locus are not identical by descent, and b and b' at a second locus are also not identical by descent. Subscripts serve to distinguish measures according to the individuals receiving the two, three or four gametes on which genes a, a', b and b' are located. For example, $\Theta_1 = \Theta_{\text{II}}$ is the probability of double nonidentity for the two pairs of genes carried on the two gametes received by individual I. Further explanation of these 10 measures is provided by WEIR, AVERY and HILL (1980). We also define P to be the panmictic index or one minus the inbreeding coefficient (probability a and a' not identical by descent in one individual) and Π one minus the co-ancestry coefficient (probability a and a' not identical by descent in two individuals).

Transition matrices for the descent measures are given for four mating systems by WEIR, AVERY and HILL (1980) following the general algorithm established by WEIR and COCKERHAM (1969). The mating systems are as follows:

MS: Monoecious with random selfing, N individuals.

ME: Monoecious with selfing excluded, N individuals.

DR: Dioecious with random pairing, each progeny from a new pairing, M males and F females. Effective population size $N_e = 4MF/(M+F)$.

DH: Dioecious with lifetime pairing to give a hierarchical structure, M males and $F = Mf$ females, each male mated to f females. Effective population size $N_e = 4MF/(M+F)$.

If v_t is the vector of descent measures $(\Theta_1, \ldots, \Delta_5)$ at generation t and $\underset{\sim}{\Omega}$ the transition matrix for that mating system, then $v_{t+1} = \underset{\sim}{\Omega} \, v_t$. The matrix $\underset{\sim}{\Omega}$ contains information about the mating system and has elements that are functions of (effective) population size and linkage parameter, λ, which is one minus twice the recombination fraction, c, between the two loci studied.

Frequencies of gene combinations: Just as there is a descent measure for each arrangement of four genes in two, three or four gametes, so there is a frequency for each arrangement when these genes have specified alleles. The frequency with which a random individual I receives two gametes both carrying alleles A and B, *i.e.*, the frequency of $AA\,BB$ double homozygotes, is written as P_{AB}^{AB}, for

TABLE 1

Definitions of descent measures and frequencies of gene combinations

Gene arrangements*	Descent measures†	Frequencies‡	$p_A p_B$	Moments $p_A p_B \times (2-p_A-p_B)$ — Coefficients of moments	$p_A(1-p_A) \times p_B(1-p_B)$ — Coefficients of moments
Digenic					
(ab)		$P_{AB}^{\,\cdot\cdot}$	1		
$(a\|b)$		$P^{A}_{\,\cdot\,B}$	1		
$(a)(b)$		$P^{\,\cdot}_{\underline{A}\,\cdot\,\underline{B}}$	1		
Trigenic					
$(ab\|a'),(ab\|b')$		$P^{AB}_{A\,.}+P^{AB}_{\,.\,B}$	2	$-P$	
$(aa')(b),(bb')(a)$		$P^{A}_{A}\|^{B}_{\,.}+P^{A}_{\,.}\|^{B}_{B}$	2	$-P$	
$(ab)(a'),(ab)(b')$		$P^{AB}_{\underline{A}\,.}+P^{AB}_{\,.\,\underline{B}}$	2	$-\Pi$	
$(a\|b)(a'),(a\|b)(b')$		$P^{A/B}_{\underline{A}\,.}+P^{A/B}_{\,.\,\underline{B}}$	2	$-\Pi$	
$(a)(a')(b),(a)(b)(b')$		$P^{A}_{\underline{A}\,.}\|^{B}+P^{A}\|^{B}_{\,.\,\underline{B}}$	2	$-\Pi$	
Quadrigenic					
Digametic					
$(ab\|a'b')$	$\Theta_1=\Theta_{\mathrm{II}}$	P^{AB}_{AB}	1	$-P$	Θ_1
$(ab)(a'b')$	$\Theta_2=\Theta_{\mathrm{II}'}$	$P^{AB}_{\underline{AB}}$	1	$-\Pi$	Θ_2
Trigametic					
$(ab)(a'\|b')$	$\Gamma_1=\Gamma_{\mathrm{I,I'I'}}$	$P^{A\ B}_{\underline{A/B}}$	1	$-\Pi$	Γ_1
$(ab\|a')(b'),(ab\|b')(a')$	$\Gamma_2=\Gamma_{\mathrm{I,\overline{II}'}}$	$P^{A\ B}_{A\|B}+P^{A\ B}_{\underline{A}\|B}$	2	$-(P+\Pi)$	$2\Gamma_2$
$(ab)(a')(b')$	$\Gamma_3=\Gamma_{\mathrm{I,I'I''}}$	$P^{A\ B}_{\underline{A\|B}}$	1	$-\Pi$	Γ_3
Quadrigametic					
$(a\|b)(a'\|b')$	$\Delta_1=\Delta_{\mathrm{II,I'I'}}$	$P^{A/B}_{A/B}$	1	$-\Pi$	Δ_1
$(a\|a')(b\|b')$	$\Delta_2=\Delta_{\mathrm{II',II'}}$	$P^{A}_{A}\|^{B}_{B}$	1	$-P$	Δ_2
$(a\|b)(a')(b')$	$\Delta_3=\Delta_{\mathrm{II,I'I''}}$	$P^{A/B}_{\underline{A\|B}}$	1	$-\Pi$	Δ_3
$(a\|a')(b)(b'),(b\|b')(a)(a')$	$\Delta_4=\Delta_{\mathrm{II',II''}}$	$P^{A}_{A}\|^{B}_{\underline{B}}+P^{A}_{\underline{A}}\|^{B}_{B}$	2	$-(P+\Pi)$	$2\Delta_4$
$(a)(a')(b)(b')$	$\Delta_5=\Delta_{\mathrm{II',I''I'''}}$	$P^{A}_{\underline{A}}\|^{B}_{\underline{B}}$	1	$-\Pi$	Δ_5

* Bars separate genes on separate gametes, parentheses separate genes in separate individuals.

† Prob(a not identical by descent to a' and b not identical by descent to b').

‡ When genes a,a' are of allelic type A and genes b,b' are of allelic type B. Dots denote summation, vertical and horizontal bars separate individuals, while diagonal bars separate gametes within individuals, as shown by the left-hand column.

example. The relation between descent measures and frequencies is simple (COCKERHAM and WEIR 1973). Providing there is no initial linkage disequilibrium this double homozygote frequency is

$$P^{AB}_{AB} = p_A p_B - p_A p_B (2-p_A-p_B) P + p_A (1-p_A) p_B (1-p_B)\,\Theta_1 \ ,$$

where p_A, p_B are the frequencies of alleles A and B. Remember that P and Θ_1 are probabilities of *nonidentity*. All frequencies required in this note are shown in Table 1, along with their relation to the descent measures. We emphasize that the frequencies are population values, or excepted values over replicate lines.

Linkage disequilibrium: Following COCKERHAM and WEIR (1977), we employ the following coefficients of linkage disequilibrium:

within individuals $\qquad D_w^{AB} = P_{..}^{AB} - P_{.B}^{A}$;

between individuals $\qquad D_b^{AB} = P_{.B}^{A} - p_A p_B$

"usual" coefficient $\qquad D^{AB} = P_{..}^{AB} - p_A p_B = D_w^{AB} + D_b^{AB}$

composite coefficient $\qquad \Delta^{AB} = P_{..}^{AB} + P_{.B}^{A} - 2p_A p_B = D_w^{AB} + 2D_b^{AB}$

and from now on drop the allelic designations. The first two coefficients are of use, or appropriate, only when coupling and repulsion double heterozygotes can be distinguished and digenic frequencies $P_{..}^{AB}$ (A, B in one gamete) and $P_{.B}^{A}$ (A, B in different gametes within one individual) can be recovered from genotypic data. If random union of gametes is invoked, of course, gametic data can be inferred from genotypic data, and in this case $D_b = 0$ and $\Delta = D = D_w$. The usual coefficient, which is a gametic measure, is seen to consist of two parts, one that serves primarily to measure departures from random union of gametes, while the composite coefficient is the only one about which inferences can be made directly from genotypic data in which coupling and repulsion double heterozygotes cannot be distinguished. It requires no assumption about the mating system.

Sampling variation of disequilibria: Observed sample values and maximum likelihood estimates of linkage disequilibria are denoted by tildes, and sampling variances of these estimates can be expressed in terms of frequencies of gene combinations.

Establishing such expressions is straightforward, although tedious, and will be sketched here for $\tilde{D}_w$. If indicator variables x_{ij}, y_{ij} are defined for the jth gamete ($j = 1,2$) in the ith ($i = 1,2,\ldots,n$) individual of a sample of n individuals as

$\qquad x_{ij} = 1$, if gamete carries A, $\qquad y_{ij} = 1$, if gamete carries B
$\qquad\quad = 0$ otherwise $\qquad\qquad\qquad\quad = 0$ otherwise

then

$$\tilde{P}_{..}^{AB} = \frac{1}{2n} \sum_i \sum_j x_{ij} y_{ij} , \qquad \tilde{P}_{.B}^{A} = \frac{1}{2n} \sum_i \sum_{j \neq j'} x_{ij} y_{ij'}$$

$$\tilde{D}_w = \frac{1}{2n} \sum_i (x_{i1} - x_{i2})(y_{i1} - y_{i2}) .$$

Recall that $\tilde{P}_{..}^{AB}$ and $\tilde{P}_{.B}^{A}$ are not directly observable from genotypic data, although their sum is observable.

Taking expectations over all possible samples of size n, and over all replicate populations

$$\mathcal{E}\, \widetilde{D}_w = P^{AB}_{\cdot\cdot} - P^{A}_{\cdot B} \;,$$

which, since we are assuming initial linkage equilibrium, is seen from Table 1 to be zero. At this stage we could, if desired, separate the two sampling processes. Suppose that $\hat{P}^{AB}_{\cdot\cdot}, \hat{P}^{A}_{\cdot B}$ were the digenic frequencies in the infinite gametic pool produced by a set of parents in a particular population. Then, taking expectations just over all samples of size n progeny from these parents would provide

$$\mathcal{E}\, \widetilde{D}_w = \hat{P}^{AB}_{\cdot\cdot} - \hat{P}^{A}_{\cdot B} \;.$$

The further expectation over all replicate populations gives population frequencies $P^{AB}_{\cdot\cdot}, P^{A}_{\cdot B}$

$$\mathcal{E}\, (\hat{P}^{AB}_{\cdot\cdot} - \hat{P}^{A}_{\cdot B}) = P^{AB}_{\cdot\cdot} - P^{A}_{\cdot B} \;,$$

which are functions of allelic frequencies p_A, p_B in the initial population from which all replicate populations are derived. These two processes will now be combined because, when predicting future behavior, particular frequencies such as $\hat{P}^{AB}_{\cdot\cdot}$ cannot be used.

Now since indicator variables are unchanged by squaring, it follows, for example, that

$$(\widetilde{D}_w)^2 = \frac{1}{4n^2} \sum_i (x_{i1}+x_{i2}-2x_{i1}x_{i2})(y_{i1}+y_{i2}-2y_{i1}y_{i2})$$

$$+ \frac{1}{4n^2} \sum_{i \neq i'}\sum (x_{i1}-x_{i2})(y_{i1}-y_{i2})(x_{i'1}-x_{i'2})(y_{i'1}-y_{i'2})$$

and

$$\mathrm{Var}(\widetilde{D}_w) = \mathcal{E}(\widetilde{D}_w)^2 = \frac{1}{2n}\left[P^{AB}_{\cdot\cdot} + P^{A}_{\cdot B} - 2(P^{AB}_{A\cdot} + P^{AB}_{\cdot B}) + 2P^{AB}_{AB}\right]$$

$$+ \frac{n-1}{n}\left(P^{AB}_{AB} - 2P^{A\ B}_{A/B} + P^{A/B}_{A/B}\right) \;. \qquad (1)$$

The effect of the mating system on variation in linkage disequilibrium is incorporated by replacing frequencies of gene combinations in results such as (1) by approprate functions of descent measures from Table 1. The results are given in Table 2. Expectations for two other quantities of interest are also given in Table 2, namely

$$\widetilde{Q} = \bar{p}_A(1-\bar{p}_A)\bar{p}_B(1-\bar{p}_B) \;, \quad \widetilde{R} = [\bar{p}_A(1-2\bar{p}_A) + \widetilde{P}^{A}_{A}]\,[\bar{p}_B(1-2\bar{p}_B) + \widetilde{P}^{B}_{B}]$$

where P^{A}_{A} and P^{B}_{B} are single homozygote frequencies. The population quantity R is just Q modified for departures from Hardy-Weinberg equilibrium and the

two quantities are equal when such equilibrium obtains. They are used (WEIR 1979) to define correlations of gene frequencies

$$r^2 = \widetilde{D}^2/\widetilde{Q}, \quad r^{*2} = \widetilde{\Delta}^2/\widetilde{R}.$$

Approximate expressions: The preceding treatment is well suited to computer iteration, but does not offer immediate insight into the behavior of the various quantities. Accordingly, we have developed some approximations that generally perform quite well and help to relate this work to previous work (AVERY and HILL 1979).

For each mating system studied here, the transition matrix can be expressed as

$$\underset{\sim}{\Omega} = \underset{\sim}{T} + \underset{\sim}{S}/N + O(N^{-2}),$$

where N is replaced by N_e for the dioecious cases, and it turns out that $\underset{\sim}{T}\,\underset{\sim}{1} = \underset{\sim}{1}$, where $\underset{\sim}{1}$ is the vector of unit elements. If the largest eigenvalue v of $\underset{\sim}{\Omega}$ and the corresponding vector $\underset{\sim}{\omega}$ are written as

$$v = 1 + r/N + O(N^{-2}), \qquad \underset{\sim}{\omega} = \underset{\sim}{1} + \underset{\sim}{x}/N + O(N^{-2}),$$

then the relation between eigenvectors and eigenvalues, $\underset{\sim}{\Omega}\,\underset{\sim}{\omega} = v\,\underset{\sim}{\omega}$, gives

$$\underset{\sim}{1} + (\underset{\sim}{S}\underset{\sim}{1} + \underset{\sim}{T}\underset{\sim}{x})/N + O(N^{-2}) = \underset{\sim}{1} + (\underset{\sim}{x} + r\underset{\sim}{1})/N + O(N^{-2})$$

or

$$(\underset{\sim}{S} - \underset{\sim}{I}r)\underset{\sim}{1} = (\underset{\sim}{I} - \underset{\sim}{T})\,\underset{\sim}{x}. \tag{2}$$

The matrix $\underset{\sim}{I} - \underset{\sim}{T}$ is not of full rank, so that (2) does not give a unique solution for $\underset{\sim}{x}$, although differences among the elements can be obtained and are displayed in Table 3. A unique value, -1, is obtained for r to give a largest eigenvalue of $1-1/N$. Notice that systems ME and DR give the same results where the effective population size N_e is set at N for the monoecious case and at $4MF/(M+F)$ for the dioecious case.

The approximations $1-1/N$ for v and $\underset{\sim}{1} + \underset{\sim}{x}/N$ for $\underset{\sim}{\omega}$ are useful, provided that N is sufficiently large that second order terms can be ignored and provided that the linkage parameter λ between the two loci is not near one, or recombination c is not near zero. When $\lambda = 1$ or $c = 0$, the largest root of $\underset{\sim}{\Omega}$ is $1-1/2N$, corresponding to the rate of fixation at a single locus since genes at both loci are now transmitted as a single unit and fixed at the same rate. The eigenvalue $1-1/N$ corresponds, for large N, to $(1-1/2N)^2 \approx 1-1/N$, which is a good approximation to the rate of loss of heterozygosity at pairs of independent loci.

The use of these approximations is seen most easily when the sample size n becomes very large, so that the variances of disequilibria are obtained for the infinite progeny array from a set of parents. In this case, Table 2 provides

$$\mathrm{Var}(\widetilde{D}_w) = (\Theta_2 - 2\Gamma_1 + \Delta_1)Q, \quad \mathrm{Var}(\widetilde{D}_b) = (\Delta_1 - 2\Delta_3 + \Delta_5)Q$$

$$\mathrm{Cov}(\widetilde{D}_w, \widetilde{D}_b) = (\Gamma_1 - \Gamma_3 - \Delta_1 + \Delta_3)Q$$

$$\mathrm{Var}(\widetilde{\Delta}) = (\Theta_2 + 2\Gamma_1 - 4\Gamma_3 + \Delta_1 - 4\Delta_3 + 4\Delta_5)Q, \quad \mathrm{Var}(\widetilde{D}) = (\Theta_2 - 2\Gamma_3 + \Delta_5)Q$$

$$\mathcal{E}(\widetilde{Q}) = (\Delta_5)Q, \qquad \mathcal{E}(\widetilde{R}) = (\Delta_2 - 4\Delta_4 + 4\Delta_5)Q,$$

TABLE 2

Variances of disequilibria in terms of descent measures (all terms to be multiplied by Q)

	Θ_1	Θ_2	Γ_1	Γ_2	Γ_3	Δ_1	Δ_2	Δ_3	Δ_4	Δ_5
$\mathcal{E}(\widetilde{D}_w)^2$	$4n^2$	$2n^2$	$-2n^2$	0	0	$2n^2$	0	0	0	0
$\mathcal{E}(\widetilde{D}_b)^2$	$(2n^2-2n+1)$	1	$-(n-1)$	$-(n-1)$	2	$(2n^2-2n+1)$	1	$-2(n-1)$	1	1
$\mathcal{E}(\widetilde{D}_w\widetilde{D}_b)$	$-2n^2$	$-n$	n^2	0	$-n$	$-n(2n-1)$	0	n	0	0
$\mathcal{E}(\widetilde{\Delta})^2$	$4(n-1)^2$	$2(n^2-2n+2)$	$2(n^2-2n+2)$	$-4(n-1)$	$-4(n-2)$	$2(n^2-2n+2)$	4	$-4(n-2)$	4	4
$\mathcal{E}(\widetilde{D})^2$	$(2n^2-2n+1)$	$(2n^2-2n+1)$	$-(n-1)$	$-(n-1)$	$-2(n-1)$	1	1	2	1	1
$\mathcal{E}(\widetilde{Q})$	1	1	1	1	2	1	1	2	1	1
$\mathcal{E}(\widetilde{R})$	$4(n-1)^2$	4	4	$-4(n-1)$	8	4	$4(n-1)^2$	8	$-4(n-1)$	4
Multiplier for column	$\dfrac{1}{4n^3}$	$\dfrac{(n-1)}{2n^3}$	$\dfrac{(n-1)}{n^3}$	$\dfrac{2(n-1)}{n^3}$	$\dfrac{(n-1)(n-2)}{n^3}$	$\dfrac{(n-1)}{2n^3}$	$\dfrac{(n-1)}{4n^3}$	$\dfrac{(n-1)(n-2)}{n^3}$	$\dfrac{(n-1)(n-2)}{n^3}$	$\dfrac{(n-1)(n-2)(n-3)}{n^3}$

TABLE 3

Elements of approximate eigenvectors corresponding to dominant eigenvalues of the descent measure transition matrices for various mating systems [deviation of x_i/N_e from x_{10}/N_e]

Element (i)	Descent measure	MS*	Mating system	
			ME and DR	DH
1	Θ_1	$\dfrac{\alpha\dagger}{N_e}$	$\dfrac{1}{N_e}+\dfrac{(1+\lambda)^2\alpha}{4N_e}$	$\dfrac{1}{N_e}+\dfrac{(1+\lambda)^2\alpha}{4N_e}+\dfrac{(1-\lambda)}{2N_e(f+1)(3+\lambda)}$
2	Θ_2	$\dfrac{\alpha}{N_e}$	$\dfrac{\alpha}{N_e}$	$\dfrac{\alpha}{N_e}+\dfrac{(1-\lambda)}{2N_e(f+1)(3+\lambda)}$
3	Γ_1	0	0	0
4	Γ_2	0	$\dfrac{1}{2N_e}$	$\dfrac{1}{2N_e}$
5	Γ_3	0	0	0
6	Δ_1	0	0	$\dfrac{1}{2N_e(f+1)}$
7	Δ_2	0	$\dfrac{1}{N_e}$	$\dfrac{1}{N_e}$
8	Δ_3	0	0	0
9	Δ_4	0	$\dfrac{1}{2N_e}$	$\dfrac{1}{2N_e}$
10	Δ_5	0	0	0

* For MS—$\Theta_1 = \Theta_2$, $\Gamma_1 = \Gamma_2 = \Gamma_3$, $\Delta_1 = \Delta_2 = \Delta_3 = \Delta_4 = \Delta_5$ *exactly.*

$\dagger\,\alpha = \dfrac{1+\lambda^2}{(1-\lambda)(3+\lambda)}$.

where $Q = p_A (1-p_A) p_B (1-p_B)$ in the base population. Notice that each of the second moments of disequilibria are contrasts in descent measures and tend to zero over time. Ratios of quantities governed by the same transition matrix can have nonzero final values, however, and expressions for the eventual correlations of gene frequencies can be obtained. Approximating the expected value of a ratio by the ratio of expected values, as is usual in this work (OHTA and KIMURA 1969), for ME, for example

$$\mathcal{E}r^2 \approx \mathcal{E}\widetilde{D}^2/\mathcal{E}\widetilde{Q} = d^2 \approx \frac{1+\lambda^2}{N_e(1-\lambda)(3+\lambda)} = \frac{c^2+(1-c)^2}{2N_ec(2-c)} , \qquad (3)$$

from Tables 2 and 3. The same value holds for $\mathcal{E}r^{*2} \approx \mathcal{E}\widetilde{\Delta}^2/\mathcal{E}\widetilde{R} = \delta^2$. We prefer to use d^2 rather than the σ^2_D of OHTA and KIMURA (1969) because we are not dealing with variances (nor, strictly, correlations). The result for d^2 has been given previously by AVERY (1978) and AVERY and HILL (1979). For the remainder of this section, we present results just in terms of $c = 1 - 2\lambda$.

Providing sample sizes n are reasonably large, so that terms of order n^{-2} can be ignored in Table 2, we can modify these approximate results to include the effects of sampling for estimation. For example

$$\text{Var}(\widetilde{D}) = \left[(\Theta_2 - 2\Gamma_3 + \Delta_5) + \frac{1}{n} \left(\frac{1}{2} \Theta_1 - 2\Theta_2 - \Gamma_1 - 2\Gamma_2 \right. \right.$$
$$\left. \left. + 8\Gamma_3 + 2\Delta_3 + \Delta_4 - 6\Delta_5 \right) \right] Q .$$

If terms involving $1/nN_e$ are also ignored, the net effect of this is to add $1/2n$ to d^2 and $1/n$ to δ^2. Tables 2 and 3 then provide, for ME, MS and DR,

$$d^2 \approx \frac{c^2 + (1-c)^2}{2N_e c (2-c)} + \frac{1}{2n} , \qquad \delta^2 \approx \frac{c^2 + (1-c)^2}{2N_e c (2-c)} + \frac{1}{n} , \qquad (4)$$

while for DH

$$d^2 \approx \frac{[(1-c)^2 + 2c^2] + f[(1-c)^2 + c^2]}{2N_e c (2-c)(f+1)} + \frac{1}{2n} ,$$

$$\delta^2 \approx \frac{(1+2c^2) + f[(1-c)^2 + c^2]}{2N_e c (2-c)(f+1)} + \frac{1}{n} . \qquad (5)$$

For $f \to \infty$, the values for DH in (5) reduce to those in (4). With monogamy, $f=1$, (5) becomes

$$d^2 \approx \frac{2(1-c)^2 + 3c^2}{4N_e c (2-c)} + \frac{1}{2n} , \qquad \delta^2 \approx \frac{1 - c + 2c^2}{2N_e c (2-c)} + \frac{1}{n} . \qquad (6)$$

Numerical values: Examples showing the magnitude of the variation in linkage disequilibrium, as indicated by d^2 and δ^2, are given for the hierarchical mating system in Table 4. Exact results are compared to the approximations (5) for a range of population and sample sizes, over which the approximations are seen to be very satisfactory.

DISCUSSION

The establishment of simple and useful approximations to the quantity d^2 has received much attention in the past, but this is the first exact treatment. WEIR and COCKERHAM (1974) did not distinguish between sample and population size, and may have obscured the simple nature of the theory by deriving transition equations for moments of the disequilibria rather than using transition arguments just for descent measures and translating these to observable quantities such as linkage disequilibrium only in the sampling generation. The earlier treatment was more general in that initial linkage equilibrium was not required, it then being necessary to use the complete set of descent measures including descent relations for nonallelic genes. The formulae derived here using the eigenvector of the transition matrix of measures hold when the number of generations becomes large, and, in that case, the initial conditions do not matter. The rate of approach to these stable values depends on the linkage parameter:

TABLE 4

Comparison of exact and approximate values of d^2, δ^2 *for DH mating system*

N_e	M	F	f	n		$\lambda = 0$	$\lambda = \dfrac{3}{4}$	$\lambda = \dfrac{15}{16}$	$\lambda = 0$	$\lambda = \dfrac{3}{4}$	$\lambda = \dfrac{15}{16}$
							d^2			δ^2	
64	32	32	1	64	Exact	0.0141	0.0329	0.1182	0.0259	0.0449	0.1309
					(Approx.)	(0.0143)	(0.0341)	(0.1272)	(0.0260)	(0.0458)	(0.1389)
				256	Exact	0.0082	0.0272	0.1126	0.0140	0.0331	0.1189
					(Approx.)	(0.0085)	(0.0283)	(0.1213)	(0.0143)	(0.0341)	(0.1272)
				∞	Exact	0.0063	0.0252	0.1108	0.0101	0.0292	0.1150
					(Approx.)	(0.0065)	(0.0263)	(0.1194)	(0.0104)	(0.0302)	(0.1233)
256	128	128	1	64	Exact	0.0095	0.0143	0.0367	0.0184	0.0233	0.0459
					(Approx.)	(0.0094)	(0.0144)	(0.0377)	(0.0182)	(0.0232)	(0.0464)
				256	Exact	0.0036	0.0084	0.0309	0.0065	0.0114	0.0339
					(Approx.)	(0.0036)	(0.0085)	(0.0318)	(0.0065)	(0.0115)	(0.0347)
				∞	Exact	0.0016	0.0065	0.0290	0.0026	0.0075	0.0300
					(Approx.)	(0.0016)	(0.0066)	(0.0298)	(0.0026)	(0.0075)	(0.0300)
256	80	320	4	64	Exact	0.0093	0.0143	0.0367	0.0177	0.0227	0.0453
					(Approx.)	(0.0092)	(0.0143)	(0.0376)	(0.0174)	(0.0226)	(0.0458)
				256	Exact	0.0034	0.0084	0.0309	0.0057	0.0108	0.0333
					(Approx.)	(0.0034)	(0.0085)	(0.0318)	(0.0057)	(0.0108)	(0.0341)
				∞	Exact	0.0014	0.0065	0.0290	0.0018	0.0068	0.0294
					(Approx.)	(0.0014)	(0.0065)	(0.0298)	(0.0018)	(0.0069)	(0.0302)
256	64	∞	∞	64	Exact	0.0091	0.0142	0.0366	0.0171	0.0223	0.0448
					(Approx.)	(0.0091)	(0.0143)	(0.0376)	(0.0169)	(0.0221)	(0.0454)
				256	Exact	0.0032	0.0084	0.0308	0.0052	0.0103	0.0328
					(Approx.)	(0.0033)	(0.0085)	(0.0318)	(0.0052)	(0.0104)	(0.0337)
				∞	Exact	0.0013	0.0064	0.0289	0.0013	0.0064	0.0289
					(Approx.)	(0.0013)	(0.0065)	(0.0298)	(0.0013)	(0.0065)	(0.0298)
1024	512	512	1	64	Exact	0.0083	0.0095	0.0152	0.0165	0.0177	0.0235
					(Approx.)	(0.0082)	(0.0095)	(0.0153)	(0.0163)	(0.0175)	(0.0233)
				256	Exact	0.0024	0.0036	0.0093	0.0046	0.0058	0.0116
					(Approx.)	(0.0024)	(0.0036)	(0.0094)	(0.0046)	(0.0058)	(0.0116)
				∞	Exact	0.0004	0.0016	0.0074	0.0007	0.0019	0.0076
					(Approx.)	(0.0004)	(0.0016)	(0.0075)	(0.0007)	(0.0019)	(0.0077)

the largest eigenvalue of $\underset{\sim}{\Omega}$ is approximately $1-1/N$ and the next largest does not exceed, but may be very close to, $(1+\lambda)/2 = 1\text{-c}$. For the *DH* mating system, the exact values shown in Table 4 were all reached to four significant figures within five generations of $\lambda = 0$ ($c = 1/2$), within 30 generations for $\lambda = 3/4$

($c = 1/8$) and between 70 and 180 generations for $\lambda = 15/16$ ($c = 1/32$). The time to equilibrium increases at an accelerating rate as linkage increases.

Our results show that the mating structure does have an effect on measures of disequilibrium such as d^2, and we find the differences to be somewhat surprising. The same formulae (4) hold for monoecious populations with or without random selfing (*MS,ME*), for dioecious populations in which pairing is random for each progeny separately (*DR*), and where there are fixed matings but very many mates for each male (*DH, f $\to \infty$*); but if there is a hierarchical structure with fixed matings and few mates per male, particularly with monogamy (*DH, f = 1*), different formulae (5,6) apply. The differences are proportionately greatest for unlinked loci ($\lambda = 0$, $c = 0.5$): for *MS, ME* and *DR*, from (4),

$$d^2 \approx 1/3N_e + 1/2n \ , \quad \delta^2 \approx 1/3N_e \, 1/n \ , \tag{7}$$

while for *DH* with monogamy, from (6),

$$d^2 \approx 5/12N_e + 1/2n \ , \quad \delta^2 \approx 2/3N_e + 1/n \ . \tag{8}$$

The difference between results (7) and (8) decreases as N_e increases and would be slight for natural populations. In a previous paper (WEIR, AVERY and HILL 1980), we considered the effects of the mating system on variances of inbreeding coefficients. The most nearly analogous expression is then the squared coefficient of variation of nonidentity, which for unlinked loci takes values of $1/3N_e$ for *MS*, $1/12N_e$ for *ME* and *DR* and for *DH* with $f \to \infty$, and $1/6 \, N_e$ for *DH* with $f = 1$. These differences could be explained in terms of the frequencies of self, full-sib and half-sib matings: with *DR*, for example, matings among full-sibs occur with frequency of order $1/N_e^2$ while with monogamy their frequency is of order $1/N_e$, but these arguments do not seem obviously relevant to the variance of linkage disequilibrium. Our present view is that the crucial difference between *MS, ME* and *DR*, on the one hand, and *DH* with monogamy on the other, is that any generation in the latter involves only one of the possible combinations of matings; in an analysis of variance we could conceptually include a component for "variance between matings." It is reasonable to assume that such a component would be nonzero here, for effective recombination occurs only in individuals heterozygous at both loci and different mating sets will lead to different frequencies of doubly heterozygous progeny. The implications of this variance between matings need to be examined in other contexts.

In the analysis, we have dealt with quantities such as the ratio of expected moments, *e.g.*, d^2 (3) rather than the expectation of the quantity that can be observed, *e.g.*, the ratio r^2. Although the analysis was carried out using diffusion-type arguments for large population sizes in which no formal account of the mating structure was taken, HILL (1977) found that r^2 was mostly a good approximation for d^2, and additional numerical results show that they agree more closely as linkage becomes loose. Thus, the differences between the mating structures derived here for d^2 with unlinked loci should hold also for r^2 and similar quantities. Other approaches (SVED and FELDMAN 1973) have dealt

directly with r^2. It then seems necessary to construct arguments conditional on observed gene frequencies.

Formulae obtained here illustrate the difficulties of using observations on linkage disequilibrium to make inferences about population size, selection and structure. In particular, LAURIE-AHLBERG and WEIR (1979) wished to use the variance of Δ, specifically $E(r^{*2})$ [equation (3)], to estimate effective population size for populations of *Drosophila melanogaster* that carried inversions, so that the analysis had to be restricted to unlinked loci. As illustrated by (7) and (8), the expectation of such quantities depends on the degree of polygamy, but more importantly, the previous population size (N_e) effects can be swamped by the final sampling (n) effects unless n is very much larger than N_e, unlikely for typical sample sizes of 100 or so.

LITERATURE CITED

AVERY, P. J., 1978   The effect of finite population size on models of linked overdominant loci. Genet. Res. **31**: 239–254.

AVERY, P. J. and W. G. HILL, 1979   Variance in quantitative traits due to linked dominant genes and variance in heterozygosity in small populations. Genetics **91**: 817–844.

COCKERHAM, C. C. and B. S. WEIR, 1973   Descent measures for two loci with some applications. Theor. Pop. Biol. **4**: 300–330. ——, 1977   Digenic descent measures for finite populations. Genet. Res. **30**: 121–147.

HILL, W. G., 1977   Correlation of gene frequencies between neutral linked genes in finite populations. Theor. Pop. Biol. **11**: 239–248.

HILL, W. G. and A. ROBERTSON, 1968   Linkage disequilibrium in finite populations. Theor. and Appl. Genet. **38**: 226–231.

LAURIE-AHLBERG, C. C. and B. S. WEIR, 1979   Allozymic variation and linkage disequilibrium in some laboratory populations of *Drosophila melanogaster*. Genetics **92**: 1295–1314.

OHTA, T. and M. KIMURA, 1969   Linkage disequilibrium at steady state determined by random genetic drift and recurrent mutation. Genetics **63**: 229–238.

SVED, J. A. and M. W. FELDMAN, 1973   Correlation and probability methods for one and two loci. Theor. Pop. Biol. **4**: 129–132.

WEIR, B. S., 1979   Inferences about linkage disequilibrium. Biometrics **35**: 235–254.

WEIR, B. S., P. J. AVERY and W. G. HILL, 1980   Effect of mating structure on variation in inbreeding. Theor. Pop. Biol. (in press).

WEIR, B. S. and C. CLARK COCKERHAM, 1969   Group inbreeding with two linked loci. Genetics **63**: 711–742. ——, 1974   Behavior of pairs of loci in finite monoecious populations. Theor. Pop. Biol. **6**: 323–354.